Appendix B | Three-Level QA Review of Coastal 2000 Northeast Database

This appendix describes the QA review process performed on Coastal 2000 data in the Northeast Region, coordinated by the Atlantic Ecology Division (AED) (U.S. EPA, 2000d). Each state or Cooperative Agreement recipient measures a suite of field data and collects water, sediment, and fish samples for laboratory analysis. The states may elect to forward the samples to a national contract laboratory or conduct the analytical analyses themselves. The results of the field and laboratory analyses are sent to AED for incorporation into a regional database. These data are subjected by AED to the three levels of QA review described below.

The states or contract laboratories provide the data in electronic form to the project officer at EPA AED. A regional database manager at the AED combines all of the states' data into a "d1-database," organized into separate data files by similarity and by states. For example, all nutrient-related data are entered into the NUTRNTS file. In turn, each data file contains several parameters; for example, the NUTRNTS file includes the nutrient parameters (e.g., nitrate. ammonium, phosphate).

The d1-database contains many parameters that are administrative in nature or descriptive of the sampling event, for example, the identity of the sampling vessel and crew and the weather conditions at the time of sampling. The AED database manager constructs a summary database, or "d2-database," consisting of parameters that have been identified to be the most useful to data users.

Level I QA Review

A Level 1 review examines the d1-database for completeness, format compatibility, and internal consistency. The checks listed below are simple and can be performed without detailed knowledge of the nature of the parameters. A Level I review is complete when all data gaps are filled or explained and obvious errors have been corrected. Records are kept of any changes made to the database. The steps for the Level I review are as follows:

(1) A completeness check is performed on all data submitted by states and laboratories. This check involves comparing the number of data entries in each file to the number of stations sampled. The database manager notes and investigates any missing data.

(2) A range check of each parameter is performed to highlight records falling outside an expected range. The database manager notes outliers and corrects any obvious errors, such as data submitted with incorrect units. Persistent outliers are highlighted for a Level 2 review.

(3) Simple consistency checks are performed by comparing independent records of closely related parameters. For instance, records of latitudes and longitudes are compared with planned locations, and water depths measured by independent methods are compared.

The AED database manager submits any questions/corrections that have been identified with suggested database changes to the Project Officer. The Project Officer transmits these questions/corrections to the Cooperative Agreement Program Manager, who resolves the concerns, concurs/non-concurs with the suggested changes, and submits a revised data file(s) if necessary. Once the Cooperative Agreement recipient concurs with the changes to the database, the Level 1 review is complete. The data files passing Level 1 QA review are made available on the password-protected Coastal 2000 Northeast Web site.

Level 2 QA Review

A Level 2 review is performed on the summary database (d2-database) parameters. The review highlights values that are unusual enough to raise the suspicions of a data user. Anomalous data include values that are especially large or small, or are noteworthy in other ways. Focus is on rare, extreme values because outliers usually merit most attention by users and may affect statistical quantities, such as averages and standard deviations.

(1) Extreme values are flagged by highlighting any record deviating from the average by more than three standard deviations.

(2) Extreme values are also highlighted visually by plotting parameter values vs station ID. The benefit of such a plot is that the outliers can be compared with nearby stations or with associated parameters. For example, if several stations in an estuary are exceptionally high or low, we would suspect that the data may be reliable. Similarly, if several closely associated parameters are extreme at a station (e.g., consistently high nutrients, or consistently high organic compounds), we would suspect that the records may be valid.

(3) Correlations among the parameters are examined. An array of miniature x-y plots is generated, one plot for each combination of associated parameters (e.g., a standard application of SAS Insight). For instance, a matrix of five water quality parameters would generate a 5x5 array of plots systematically varying in variables for the x- and y-axes. Typical plots show a regular relationship between the plotted parameters. Anomalous data are readily evident on these plots. Examination of closely related parameters may resolve questions regarding the accuracy of anomalous data.

Documentation of suspicious data identified is prepared, with invalid data flagged. This documentation becomes part of the metadata. Level 2 data are made available on the same Web site as the Level 1 data.

Level 3 QA Review

A Level 3 review is conducted to evaluate whether data submitted by the states or laboratories are comparable across areas, recognizing that the magnitudes of the values may indeed be different in the various geographic areas.

(1) A regional map is prepared for each measured parameter. Discrete map symbols denote station location and the magnitude of the parameter (e.g., low, moderate, or high). The maps are examined for noteworthy patterns that may be attributed to database errors.

(2) A bar chart is prepared for each measured parameter. The chart shows the percent area of each state's waters designated by a condition category (e.g., low, moderate, or high). The charts are also examined for anomalous patterns that may indicate database irregularities.

(3) A distribution graph is prepared for each parameter, grouping data by estuarine system to compare the range and distribution of measured values across the states.

(4) A table is prepared for each parameter summarizing the descriptive statistics of parameters by state. Although the magnitude of a parameter may vary by state, it is expected that the coefficient of variation should be roughly equivalent across the states.

A summary report is prepared, utilizing the maps, charts, and tables developed in the Level 3 review. This report is made available on the same Web site that the Level 1 and Level 2 data are available on.

Records are maintained of all data files examined and entries considered anomalous. The Project Officer reports the anomalies to the Cooperative Agreement recipient or contract laboratory data managers, who correct and resubmit the data. All changes to the original database are documented.